Perspectives in Science (2016) 8, 93—95

ScienceDirect

journal homepage: www.elsevier.com/pisc

Computational neural network regression model for Host based Intrusion Detection System☆

Sunil Kumar Gautam*, **Hari Om**

Department of Computer Science & Engineering, Indian School of Mines, Dhanbad, India

Received 9 February 2016; accepted 11 April 2016
Available online 20 April 2016

KEYWORDS
Intrusion detection;
Generalized Regression Neural Network;
Multilayer Perceptron Neural Network;
Confusion matrix

Summary The current scenario of information gathering and storing in secure system is a challenging task due to increasing cyber-attacks. There exists computational neural network techniques designed for intrusion detection system, which provide security to single machine and entire network's machine. In this paper, we have used two types of computational neural network models, namely, Generalized Regression Neural Network (GRNN) model and Multilayer Perceptron Neural Network (MPNN) model for Host based Intrusion Detection System using log files that are generated by a single personal computer. The simulation results show correctly classified percentage of normal and abnormal (intrusion) class using confusion matrix. On the basis of results and discussion, we found that the Host based Intrusion Systems Model (HISM) significantly improved the detection accuracy while retaining minimum false alarm rate.

Introduction

In recent years, computer network security is a major concern of the computer society due to the development of technologies and internet services at rapid pace. Intrusion Detection System (IDS) monitors the single machine or computer network for intruder. In 1983, Dorothy developed the first model for intrusion detection. This Intrusion Detection Expert System (IDES) analyses audit records of abnormal pattern of the system. Generally, the IDS is divided into two types of categories, namely: Network based Intrusion Detection System (NIDS), which monitors network traffic by packet sniffing technique and detects malicious activity in this traffic, and Host based Intrusion Detection System (HIDS), which monitors or analyses system log files and detects malicious or intrusive activities in a single machine. IDS uses two types of approaches, viz. misuse based and anomaly based, for intrusion detection. The former approach detects a particular pattern of attack. In contrast, the latter approach detects both specific pattern of attacks as well as new pattern of intruders (Denning, 1987; Deepa and Kavitha,

☆ This article belongs to the special issue on Engineering and Material Sciences.

* Corresponding author. Tel.: +91 7677427797.
E-mail addresses: gautamsunil.cmri@gmail.com (S.K. Gautam), hariom4india@gmail.com (H. Om).

http://dx.doi.org/10.1016/j.pisc.2016.04.005

2012). Researchers have developed several data mining approaches for intrusion detection, including *k*-Means, fuzzy logic, Genetic Algorithm, Neural Network, Support Vector Machine, etc. (Agrawal and Jitendra, 2015). In this paper, we discuss about Generalized Regression Neural Network (GRNN) model and Multilayer Perceptron Neural Network (MPNN) model for HIDS. The rest of the paper is organised as follows: The succeeding section discusses related works, with the subsequent one describing the proposed neural network approach. This is followed by the fourth and fifth sections that present the experimental methodology and confusion matrix, respectively. Next, we discuss about the simulation results and analysis. Finally, in the last section, we draw the conclusions.

Related works

In recent years, the IDS is a reliable system, which effectively monitors the networks and detects the various network attacks. Many researchers have designed several IDS using different fields, such as (give some 2—3 field names followed by etc.). The Data Mining field effectively removes false positive alarm (Al-Mamory et al., 2008). In the Data Mining field, a neural network technique most effectively identifies and forecasts abnormal activities. In this technique, different neural network algorithms, such as feed forward network, back propagation neural network, probabilistic Boolean network, etc., are used for intrusion detection. These algorithms divide dataset into training and testing datasets (Shun and Malki, 2008). Haddadi et al. have proposed IDS using feed-forward neural network with back propagation algorithm for network based intrusion detection. This scheme has used KDD-CUP'99 dataset for classification of network attacks (Haddadi et al., 2010). Neural Network techniques also increased intrusion detection system performance using DARPA dataset. Kumar et al. (2014) have used artificial neural network technique in anomaly detection and the simulation results demonstrated that there is significant improvement in anomaly detection accuracy. Abou Haidar et al. (2015) have proposed anomaly based detection system using multilayer perceptron and self-organizing maps technique to improve reliability of intrusion system.

Proposed Neural Network Model

The neural network techniques have been broadly used for intrusion detection since these techniques does not require more parameters for optimization of results. In neural network, initially N samples were given as input and it predicts the behaviour of $(N+1)$th sample using these first N samples, which is output of neural network (Salmasi et al., 2011).

This paper essentially deals with discussion about Generalized Regression Neural Network (GRNN) model and Multilayer Perceptron Neural Network (MPNN) model for host based intrusion systems using log files, which are generated by a single personal computer.

Generalized Regression Neural Networks (GRNN)

The Generalized Regression Neural Network (GRNN) approach is used for prediction of system, i.e. system behaves normally or abnormally. This technique is a fast learning algorithm, which takes all numbers of sample data to the optimal regression surface. GRNN is most useful for unplaced data sample since the regression surface is instantly defined with one data sample. GRNN approach contains four layers, namely: Input layer, Hidden layer, Pattern layer and Decision layer. At first, Input layer takes all Input values for expected target output and send it to Hidden layer. In Hidden layer, each neuron contains training dataset, and after processing it, the resulted value is sent to Pattern layer. The Pattern layer contains all target values summation of neurons. The Decision layer contains the result of predicates target value (Specht, 1991; Devaraju et al., 2014).

Multilayer Perceptron Neural Network (MPNN)

The Multilayer Perceptron Neural Network technique is also known as multi-layer neural network. MLP is classified into Feed Forward Neural Network (FFNN) and Back Propagation Neural Network (BPNN). This technique contains three layers for implementation of namely, input layer, hidden layer and output layer. In FFNN phase, all parameters are fixed and the error is computed using the formula,

$$e_i = d_i - y_i \tag{1}$$

here d_i represents desired response, y_i is the actual output, which is produced by network input. In back propagation phase, e_i is the error signal propagated by the network (Devaraju et al., 2014).

Experimental methodology

In this paper, we have taken offline dataset generated by the personal computer for evaluating the performance of GRNN and MPNN techniques. In this dataset, each data point corresponds to either normal or abnormal class. The abnormal data points are assumed as intruder data, which are generated while disabling some driver of personal computer, including USB driver, Audio driver, Graphic driver, etc. In this paper, we have taken 15 randomly selected features from the log file that contains 20,000 records. The configuration of our personal computer is given below:

- Intel(R) Core (TM) i5
- 4.00 GB RAM
- Microsoft Windows8
- 64-bit Operating System

Confusion matrix

In this paper, we have used confusion matrix to compute the intrusion detection accuracy of GRNN and MPNN models for host based system (Godbole, 2002). Generally, the confusion matrix technique is used for classification problem since it clearly represents actual class and predicated

Table 1 Confusion matrix.

Actual class	Predicated class	
	Anomaly class	Normal class
Anomaly class	True negative (TN)	False positive (FP)
Normal class	False negative (FN)	True positive (TP)

TP – Number of data points classified as Normal while they actually were Normal.
TN – Number of data points classified as Intrusion while they actually were Intrusion.
FP – Number of data points classified as Intrusion while they actually were Normal.
FN – Number of data points classified as Normal while they actually were Intrusion.

Table 2 Analysis result in percentage.

	GRNN	MPNN
Detection accuracy	98.46	97.68
Recall	97.22	96.12
Precision	95.73	97.65

class output. It determines pairwise resemblance between all sets of classes using some resemblance metrics. Table 1 represents the confusion matrix.

$$\text{Detection Accuracy} = \frac{(TN + TP)}{(TN + TP + FN + FP)} \quad (2)$$

$$\text{Precission} = \frac{(TP)}{(TP + FP)} \quad (3)$$

$$\text{Recall} = \frac{(TP)}{(TP + FN)} \quad (4)$$

Results and discussion

In this study, an attempt has been made to use GRNN and MPNN data mining models for HIDS on personal computer. We have taken some previous days of training and testing data from log files, which are stored in our personal machine in comma separated values (.csv) format for experimental analysis. The dataset contains 20000 records, which are described by 15 dataset features. The experimental results have been summarized in Table 2. We can observe from Table 2 that the intrusion detection accuracy of GRNN is 98.46. It may be noted that the recall value represents the sensitivity of our model, i.e. correctly classified abnormal behaviour of the system. The precision values defines the true intrusion in our system among normal and abnormal class.

Conclusions

This paper has discussed about the two most popular Data Mining host based GRNN and MPNN network models. The performance of GRNN and MPNN models in HIDSs has been evaluated using the log files generated on a personal machine. From the analysis performed on our datasets, we found that GRNN outperforms GRNN outperforms in terms of system accuracy rate and recall value, but the precision value is less than MPNN. Based on the simulation results, we can conclude that both GRNN and MPNN models are suitable for host based intrusion. In future, we will extend our work for online intrusion detection and may also apply different types of Data Mining techniques for offline intrusion.

References

Abou Haidar, G., et al.,2015. High perception intrusion detection system using neural networks. In: CISIS. IEEE, pp. 497–501.

Agrawal, S., Jitendra, A., 2015. Survey on anomaly detection using data mining techniques. Procedia Comput. Sci. 60, 708–713.

Al-Mamory, S.O., et al., 2008. IDS alarms reduction using data mining. In: Neural Networks. IEEE, pp. 3564–3570.

Deepa, A.J., Kavitha, V., 2012. A comprehensive survey on approaches to intrusion detection system. Procedia Eng. 38, 2063–2069.

Denning, D.E., 1987. An intrusion-detection model. IEEE Trans. Softw. Eng. 2, 222–232.

Devaraju, S., et al., 2014. Performance comparison for intrusion detection system using neural network with KDD dataset. ICTACT J. Soft Comput. 4 (3), 743–752.

Godbole, S., 2002. Exploiting confusion matrices for automatic generation of topic hierarchies and scaling up multi-way classifiers. In: Annual Progress Report. Indian Institute of Technology, Bombay, India.

Haddadi, F., et al.,2010. Intrusion detection and attack classification using feed-forward neural network. In: Computer and Network Technology (ICCNT). IEEE, pp. 262–266.

Kumar, S., et al.,2014. Increasing performance of intrusion detection system using neural network. In: ICACCCT. IEEE, pp. 546–550.

Salmasi, M., et al.,2011. Comparison of multilayer perceptron and generalized regression neural networks in active noise control. In: PACCS. IEEE, pp. 1–4.

Shun, J., Malki, H.A.,2008. Network intrusion detection system using neural networks. In: Natural Computation, ICNC'08, vol. 5. IEEE, pp. 242–246.

Specht, D.F., 1991. A general regression neural network. IEEE Trans. Neural Netw. 2 (6), 568–576.